

СТАТИСТИЧНИЙ СПОСІБ ЗГОРТАННЯ НАВЧАЛЬНОГО НАУКОВОГО ТЕКСТУ.

Шокуров О.В.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків*

Комп'ютеризація суспільства дозволяє не тільки підвищувати комфортність і якість нашого побуту, а й по-новому підходити до вирішення деяких проблем.

Однією з таких проблем є проблема навчання роботі іноземних студентів з науковим текстом, не тільки його розумінню, а й умінню його згортання. Навчання згортання тексту до послідовності ключових слів (КС) є одним з пріоритетів як шкільного, так і вищівського освітніх процесів. КС – це ті лексичні елементи тексту, за допомогою яких може бути виражений зміст, основна інформація цього тексту.

Останнім часом КС стали важливим елементом SEO-оптимізації тексту (пошукової оптимізації), оскільки правильно визначені КС є важливими для коректної роботи пошукових систем внаслідок високої частоти їх повторів в тексті. У російськомовному сегменті інтернету з'явилася значна кількість сторінок, де аналізується текстова інформація сайтів для покращення їх позицій в пошукових системах з метою підвищення їх конкурентоспроможності.

Багато сайтів аналітико-синтетичної обробки тексту пропонують пошук не тільки КС, а й т.зв. семантичного ядра текстової інформації сайту. Семантичне ядро - це речення або словосполучення, в якому є два або більше КС.

Найбільше КС знаходиться в заголовку наукового тексту. Це пояснюється важливою функцією заголовка наукового тексту. Саме в заголовку найчастіше виражена основна ідея, тема або рема тексту.

Таким чином, після згортання лінгвістичним способом тексту можна надати студентам можливість перевірити результати цієї роботи статистичним (математичним) способом згортання наукового тексту. Алгоритм цього способу складається з наступних дій: проведення статистичного аналізу лексичного складу тексту, знаходження найбільш частотних слів, представлених в заголовку тексту або в семантичному його ядрі, виписуванні речень (бажано ключових речень) з цими найбільш частотними словами в порядку, відповідному логіці тексту.

В результаті отримуємо послідовність тез, в яких відбивається основний зміст всього тексту.

Статистичний аналіз тексту (або SEO-аналіз) можна зробити на будь-якому з численних сайтів, що пропонують цю послугу. Якщо ж є проблеми з виходом в інтернет, можна скористатися однією з програм для статистичного контент-аналізу текстів. Таких програм досить багато. Наприклад, textstat, Yoshikoder, wordstat і т.п. Ми зупинили свій вибір на безкоштовній і досить зручній програмі Textus Pro. Ця програма дозволяє проводити аналіз тексту, використовуючи достатній набір налаштувань.